Fermi National Accelerator Laboratory

FERMILAB-Pub-98/034-A

The Sloan Digital Sky Survey – Pi on the Sky

Heidi Jo Newberg

Fermi National Accelerator Laboratory
P.O. Box 500, Batavia, Illinois 60510

January 1998

Submitted to *The Beam Line*

Operated by Universities Research Association Inc. under Contract No. DE-AC02-76CH03000 with the United States Department of Energy

The Sloan Digital Sky Survey - Pi on the Sky

Heidi Jo Newberg

January 28, 1998

1

When the originators of the Sloan Digital Sky Survey (SDSS) met at O'Hare International airport in the fall of 1988, their intent was to form a collaboration that would measure the size of the largest structures of galaxies in the universe. Previous galaxy surveys had shown that the largest structures were at least 400 million light years in extent - as large as the largest structures that *could* have been found by these surveys. In particular, the results of the CfA Redshift Survey were astounding. From the spectra of one thousand galaxies, the researchers were able to depict a slice of our universe with large "voids" (where the galaxy density was very low) surrounded by dense "walls" of galaxies.

Cosmology had already seen the demise of the *perfect cosmological principle* in 1929 with Edwin Hubble's discovery that the universe is expanding, and therefore changes over time. We continue to believe, however, in the *cosmological principle* - that the universe is homogeneous and isotropic. To demonstrate the validity of this basic assumption about our universe, we must be able to find some volume of the universe that is representative of the whole. It is clear that a much larger galaxy redshift survey is required to find a representative sample of the universe.

With projects to measure ten to one hundred thousand galaxies already planned or underway, the early organizers of the Sloan Digital Sky Survey proposed a redshift survey of one million galaxies. Rather than looking at a slice of the universe, this new survey would measure the position of every galaxy in a patch of sky π steradians (one quarter of the sky) in size. The survey would measure the distances to galaxies three magnitudes fainter (about four times farther away) than those observed in the CfA Redshift Survey. In addition, these million galaxies would not be chosen from the photographic sky surveys already in existence. Rather, they would be selected from a new, carefully controlled survey of the sky using a large CCD camera. By imaging the sky in several optical passbands, including an ultra-violet passband, the data from this sky survey would also be used to select quasi-stellar objects (QSOs). As the most distant collapsed objects ever observed in our universe, these would give us information about the structure of the universe on the largest possible scales. The survey was projected to take five years to build, with an additional five years of operation to complete the scientific objectives.

Although it may not have been recognized at the time, the addition of the imaging survey transformed the SDSS project from an ambitious attempt to trace the large scale structures in the universe into a plan to statistically sample *everything* in a large corner of the visible universe. What these planners had dreamed up was an imaging survey covering ten thousand square degrees of the sky in four filters; a catalog of the 70 million stars, 50 million galaxies, and one million QSOs visible in the imaging survey; and a spectroscopic survey of more than a million of these objects; all rolled into one enormous project. This statistical sample will have a tremendous impact not only on our understanding of the largest structures, but on every aspect of astronomy.

Any one of these three projects (the imaging survey, the catalog, or the spectroscopic survey) would have been considered large by the standards of ground-based astronomy. Any one of the three could be scientifically justified on its own merit. All together, the project is as colossal as its impact will be on astronomy. Okay - the goals, the time-line and the budget were optimistic. But if we were not attempting the impossible, we would not be on the forefront of research.

2

The Sloan Digital Sky Survey has attracted the active participation of over one hundred scientists, engineers, and software professionals from eight astronomy groups and departments, including: Princeton University, The University of Chicago, The Johns Hopkins University, the Japan Promotion Group (scientists at the Universities of Tokyo and Kyoto), the United States Naval Observatory, the University of Washington, the Institute for Advanced Study, and Fermi National Accelerator Laboratory. The survey is being carried out under the auspices of the Astrophysical Research Consortium (ARC) and has received significant funding, totaling about 54 million dollars, from the Alfred P. Sloan Foundation (New York), from the National Science Foundation, and from each of the member institutions.

The goals and scope of the project have changed only slightly from those put forth by the "O'Hare group." Since the main survey area is not observable during part of the year, three extra strips of sky have been added to fill in the gap. Also, we have added one passband to the imaging survey, for a total of five filters. Mostly, we have made tremendous progress designing and building the hardware and software necessary to assure our success.

The astronomical data for the SDSS will be obtained from two dedicated telescopes located at Apache Point Observatory in Sunspot, New Mexico. The data will be partially processed at the observatory before being sent to Fermi National Accelerator Laboratory, where the majority of the data processing, storage, and distribution will take place. The main SDSS telescope has a primary mirror 2.5 meters in diameter and a field of view three degrees in diameter. It will support two instruments: a photometric camera containing 54 CCDs and a spectrograph with 640 fibers. A fully automated 24-inch diameter telescope will operate simultaneously. The survey software is designed to operate these telescopes, plan imaging and spectroscopic observations so as to minimize the survey time-to-completion, acquire the data from all survey instruments, process imaging data into catalogs of astronomical objects and their associated parameters, calibrate the positions and luminosities of the measured objects, merge the data from different CCDs and different nights into one large catalog, select from this catalog the sources for which we will obtain spectra, organize the targeted objects into separate spectroscopic exposures, reduce spectroscopic exposures into lists of objects with classifications and redshifts, and store the results of all of these steps in a large database.

3

The SDSS is aggressively charting new territory both in the design of the telescope and instruments, and in the processing and acquisition of the scientific data. The spectrograph will be capable of observing more objects at one time than any other in the world. The photometric camera will have more pixels in the focal plane than any other CCD camera in existence. Our catalog of objects will be the largest, and will have better positional and photometric accuracy than any other catalog of its kind. In order to assure the astrometric and photometric uniformity we require for describing our statistical samples of the sky, we have included in the design several novel instruments which will allow us to evaluate and calibrate the data better than any previous survey. I will discuss here only a few of the innovations which make the survey possible.

The 2.5 meter telescope is specially designed to reduce "dome seeing," the distortion of images caused by turbulence in the air very close to the telescope. To reduce distortion caused by disruption of the laminar flow of air over the observatory, the SDSS telescope is cantilevered over the edge of a cliff in the direction of the prevailing wind. In addition, we use a roll-off building which eliminates the telescope building as a potential cause of heat, which also contributes to image distortion. During operation, the telescope is protected from wind and stray light by a baffle that is mechanically separated from the telescope, but that moves and tracks with it.

In all areas of optical astronomy except surveys, data from CCD cameras has supplanted data from photographic plates. CCD cameras, unlike the plates, have linear response functions and much higher efficiency for detecting light, which makes possible more accurate photometric (luminosity) measurements. Until now, these cameras were not used for surveys because it was not possible to cover enough sky with one camera for a large area survey to be tractable. By building a camera with 30 large CCDs and using a drift-scanning technique, we will not only be able to survey large areas of sky, but also to obtain the images simultaneously in five filters.

In addition to the array of 30 photometric CCDs, the focal plane contains 22 smaller astrometric CCDs, which are used to calibrate the positions of the survey objects. These CCDs image both the astrometric standard stars (which are saturated on the photometric CCDs) and also some of the brighter stars that will be unsaturated on the photometric CCDs. This allows us to tie our data to a coordinate system that is fixed on the sky. Also, it allows us to more accurately measure the relative positions of the objects found in separate CCDs in the photometric array. By comparing the same stars as imaged at the beginning and end of the CCD array, we can assess how well the telescope is tracking its trajectory in the sky.

The photometric accuracy of our catalogs will be limited by our ability to characterize the atmospheric conditions during the night. Even though photometry will be attempted only on the clearest, most stable nights, we will measure the transparency of the atmosphere as a function of time, filter, and position in the sky. The suitability of a given night for photometry will be determined with data from a weather station which logs the temperature, wind speed and direction, humidity, and dust level. In addition, the weather station includes a camera which images the whole sky at 10 microns every 20 minutes. At this wavelength, clouds stand out very clearly against the dark sky. This unique camera alerts us to weather changes during the night, and also detects lone clouds on what is otherwise a completely clear night.

The 24-inch "monitor" telescope has three functions in the survey. First, it will be used to calibrate bright stars of known luminosity with the specially designed SDSS filters. Second, the telescope will allow us to use these bright primary photometric standards to calibrate a fainter and more numerous set of secondary photometric standard stars. These secondary standards will be unsaturated in the photometric array, and allow us to directly calibrate our scans of the sky. Last, this telescope will repeatedly image the primary standard stars throughout the night to track the transparency of the atmosphere. The use of a separate telescope to track the atmosphere is unprecedented in sky surveys.

4

With all of these innovations in hardware, the software is required to come up with innovative solutions to process the data coming in. Some of the challenges are unique to our survey, such as separating blended objects and matching up the detections in all five filters, merging catalogs of objects measured in separate strips of sky, optimizing the placement of the spectroscopic exposures, and planning observations to reduce time-to-completion. Even tasks that have been done many times before, such as removing instrumental effects from the raw data, object detection and measurement, and extraction and calibration of spectral data must be optimized for our experiment. In addition, the data must be analyzed within a week or two after it is obtained so that spectrocopic targets can be selected for observation during the next lunation.

By far the most critical challenge for the software is the ability to produce highly accurate and uniform results without human intervention. In one night, the imaging camera for this survey will write to tape 140 gigabytes of data at a rate of 4.8 megabytes per second. By the time the survey has finished, it will have generated on the order of 20 terabytes of imaging data, organized into 3.3 million 6-megabyte images. It would take four and a half years of solid 40 hour weeks for one person to devote ten seconds to each of these images. It would take an additional 1.5 person-years to spend ten seconds on each spectrum. In practice, we will be able to look critically at only a tiny fraction of the data. This means the software will have to identify and correctly handle the brightest stars, the faintest galaxies, satellite trails, globular clusters, nebulae, cosmic ray hits, telescope glints and reflections, diffraction spikes, and big spiral galaxies *on its own.* If the image processing software de-blends a big galaxy into many smaller objects, or finds many objects along a satellite trail, the final catalog of objects will be contaminated. Then, the processes that automatically calibrate the data and choose spectroscopic targets will have to be clever about discarding questionable data to prevent large sections of the catalog from being mis-calibrated, and from wasting many spectroscopic observations on junk. The uniform results rely not only on the software that identifies the objects, but also on the software that records the time, conditions, and telescope position when the data was acquired; the software that controls the automated monitor telescope and reduces the images; the software that selects several dozen different types of spectroscopic targets with their individual selection criteria; and on our ability to effectively monitor the process.

Building the data processing and storage systems necessary to run the SDSS requires a higher level of infrastructure than is available at any of the universities involved with this project. In implementing the data acquisition system, the infrastructure for the data processing software, and the mechanisms for data storage, we have benefited from Fermilab's many years of experience with high energy physics experiments. Like high energy physics experiments, the scientific objectives, instruments, and software are provided by scientists at Fermilab and at each of the institutions in the collaboration. The staff at Fermilab supplies the expertise in managing large scientific projects, and the accompanying infrastructure that bring the project together. Fermilab staff have been instrumental in instituting coding standards; maintaining code management and versioning systems; specifying and procuring hardware; supporting the database; and making vast amounts of storage on tape robots available to the collaboration. Because of our presence at Fermilab, we were one of the first projects in the world to put our documentation on the World-Wide-Web (we were there before X-Mosaic, let alone Netscape, ever existed!). In the future, we will be using Fermilab's resources and experience to operate our production system. In return, experience gained from implementing the new technologies used by our project is being used to benefit high energy physics projects of the future.

5 Drift-Scanning the Sky (sidebar)

The SDSS camera will drift-scan the sky, rather than using the more common point-and-shoot method, in order to increase the fraction of the time that the imaging camera is integrating light from the sky and to reduce the number of images that must be pieced together. Point-and-shoot observations are obtained by tracking the apparent motion of the target object in the sky, and opening and closing a shutter to expose the detector. Drift-scanning is usually done by fixing the position of the telescope and moving the photo-sensitive material to track the target object. The Sloan Digital Sky Survey will have to move both the telescope and the detector to image the survey area.

Before explaining how the SDSS camera works, let me first describe a much simpler drift-scanning camera. Imagine a telescope in a fixed position on the Earth's equator, and pointing directly overhead. The telescope focuses light from the sky onto a single CCD in the focal plane. As the Earth turns, stars will enter the field-of-view of the CCD, travel across it at constant speed, and then disappear from view. To drift-scan, the two dimensional array of pixel detectors on the CCD must be aligned so that the crossing star travels exactly along one column. As light from the star moves from one row of the CCD to the next, all of the accumulated photoelectrons are also moved to the next row. This leaves one empty row of pixels at the beginning of the CCD, ready to start exposing a new part of the sky. The photoelectrons in the last row are read out and digitally stored in a computer. The camera accumulates data continuously along one equatorial strip of sky, without stopping to read the data out of the CCD while the shutter is closed. The effective exposure time of the data is the crossing time of a star across the camera.

To scan across the sky in a direction that is not along constant latitude or that is far from the equator requires the telescope to track and (in most cases) the CCD to rotate. Tracking is also required to drift at arbitrary rate (which sets the exposure time) across the sky. The first telescope and CCD combination capable of driven drift scans in arbitrary directions was the Fermilab Drift-Scan Camera mounted on the ARC 3.5 meter telescope at Apache Point Observatory, adjacent to the SDSS telescope site. The camera, which was commissioned in 1994, was built as a prototype for testing SDSS data acquisition software.

The SDSS camera puts 30 large, 2048×2048 pixel CCDs in the focal plane of the telescope - six columns of five CCDs. The six columns each scan a separate strip of sky while the camera is imaging. Each of the five CCDs in a given column images the same strip of sky, but using a different filter. The effective exposure time in each filter is about 55 seconds. A given object will first traverse the CCD with the r' filter, then the i', u', z' and g' filters, in succession. It takes about 5.8 minutes to traverse all five filters. Each column images a strip of sky 13.7 arc minutes wide which increases in length by 15 degrees per hour. A similar camera operated in point-and-shoot mode would spend about as much time exposing a 55 second image as it did reading out the CCD. Also, we would need to piece together 250,000 individual images rather than about 1000 long, continuous strips.

In addition to the photometric CCDs, there are 22 astrometric chips and two focus chips in the focal plane. Since these CCDs are the same width and pixel scale as the photometric chips, each row is read out with the same frequency - producing the same 9.5 megabytes per minute per CCD. Fewer rows in the astrometric and focus CCDs produce shorter exposure times rather than lower data rates. The data from the focus chips will be used to automatically adjust the focus in real time. Twelve of the astrometric chips, those at the leading and trailing edge of each column of photometric CCDs, will be used to assure that the rotation of the camera is aligned with the transit of the sky across the camera and to measure the uniformity of the tracking rate. The ten interleaving astrometric CCDs tie together the positions of the objects found in adjacent columns of CCDs.

The SDSS photometric camera will image 164 square degrees of sky on an average night. Including overhead and overlaps between strips of sky, we will be able to cover the SDSS survey area in ninety dark, photometric nights.